图书情报工作
第 65 卷 第 24 期　2021 年 12 月

数字人文项目发展的特色和建议*
——基于对国际数字人文获奖项目的评析

■ 单蓉蓉[1]　刘炜[2]　陈涛[3]　李惠[4]

[1] 上海大学图书情报档案系　上海 200444　[2] 上海图书馆上海科学技术情报研究所　上海 200031

[3] 中山大学信息管理学院　广州 510006　[4] 南京农业大学人文与社会发展学院　南京 210032

摘　要：[目的/意义]"数字人文奖"是国际数字人文领域的权威奖项之一，通过对获奖项目的梳理分析，可以发现数字人文领域"优秀项目"和"实用工具"的共性，总结失败项目的不足。[方法/过程]利用文献调研法、内容分析法和信息计量梳理"数字人文奖"历年获奖项目，归纳总结获奖项目的特点和不足。[结果/结论]提出数字人文的工具开发应以代码开源和用户需求为导向；项目实践应以关联数据、国际图像互操作框架(IIIF)、资源描述框架(RDF)和人工智能技术(AI)为主要技术框架；资金或资源受限的项目可选择"最小计算"为实践模式。

关键词：数字人文奖　工具开发　技术实现　最小计算

分类号：G250.7

DOI：10.13266/j.issn.0252-3116.2021.24.002

近年来数字人文的应用实践如火如荼，有以方志物产知识库、中文古籍循证平台为代表的数据基础设施建设；有深度学习模型、文本挖掘技术、知识图谱、GIS 等技术在语料库的应用；有对以"欧洲时光机"为代表的欧美地区数字人文实践的案例分析；有对如"数字敦煌""上海图书馆历史文献平台"等文化遗产资源的项目开发。学者们习惯从文献计量角度，宏观分析数字人文领域的应用热点，他们有以下共识：语言学、艺术学和史学是数字人文的主要应用领域；数据基础设施的建设是目前数字人文实践的热点之一[1-2]。尽管有学者认为当今数字人文的发展偏向应用与实践[3]，但是对数字人文项目实践的评价和总结却很少。

奖项是文化生产领域内一种重要的确认形式，是对领域内最好的生产实践作品的认可和奖励，多元奖项的认定更是能够推动学科共同体的建设。在颁奖相关活动的背后，是大量资料的收集和数据库的建立，如果把每年入围和获奖作品资料进行比较查看，可作为该领域实践发展的一种"物证"。因此本文从"数字人文奖"的历年获奖项目入手，对数字人文领域的应用实践进行分析总结，从中窥探现状和探讨趋势。

1　相关实践调研

20 世纪 90 年代，信息技术和人文学科的互动下诞生的新兴学术领域"数字人文"逐渐成为一个独立学科。以国际数字人文组织联盟(The Alliance of Digital Humanities Organizations，ADHO)为代表的学会和研究机构为推进数字人文发展，设置了很多奖项如表 1 所示。影响力较为广泛的有：ADHO 以数字人文先驱者罗伯托·布萨命名的布萨奖(Roberto Busa Prize)，被视为数字人文领域的最高个人成就奖；美国人文基金会 NEH 设立的"数字人文进步奖"(Digital Humanities Advancement Grants，DHAG)，主要资助数字人文项目的实施；加拿大数字人文学会(Canadian Society for Digital Humanities，CSDH/SCHN)面向本国学者设立的一系列奖项，如"杰出贡献奖"(Outstanding Contribution Award)是对加拿大籍的研究人员在数字人文领域的重大贡献的认可；"数字人文奖"(Digital Humanities Awards：Highlighting Resources in Digital Humanities，DH Awards)是一个开放、多元的奖，以表彰全球数字人文的资源、工具和专业知识；美国历史协会颁发的罗伊·

* 本文系国家社会科学基金项目"徽州民间文书抢救性保护与数据库建设研究"(项目编号：19BTQ010)研究成果之一。

作者简介：单蓉蓉，馆员，博士研究生，E-mail：shan_rongrong@163.com；刘炜，副馆(所)长，研究员，博士，博士生导师；陈涛，副教授，博士；李惠，副教授，博士。

收稿日期：2021-05-06　修回日期：2021-09-14　本文起止页码：18-27　本文责任编辑：易飞

罗森茨威格数字历史创新奖(The Roy Rosenzweig Prize for Innovation in Digital History)旨在奖励在数字历史领域内免费、创新的新媒体项目。

“数字人文”作为一个学科名称和专业术语的历程并不长，因此对该领域的奖项研究尚处空白阶段。调研整理的奖项如表1所示，大部分奖项以表彰研究人员或对项目提供基金资助为主，面向项目的奖项或是局限在特定学科、机构内，或是持续时间短，不能完全反映数字人文实践的历史发展。其次，这些奖项是由专家组成的评委会进行评审，专委会的评审制度虽然保证奖项的专业性和权威性，但受众在认可数字人文的学者范围内，没有积极吸引传统人文学者的参与，达不到“破圈”的效果。再者，表1中大部分奖项的设置较为单一，围绕人文资源进行的数据库建设、程序设计、可视化呈现、数字期刊和博客发布并非传统的学术活动和成果，目前被认定或转换为现有学术评价体系接受的学术成果较为困难，作为新兴学科的数字人文需要多元奖项对其成果的认可和宣传，吸引更多的关注。

表1　数字人文领域奖项列表

奖项	起始年份	评选对象	说明
罗伯托·布萨奖(Roberto Busa Prize)[4]	2002年至今(每三年)	有杰出贡献的数字人文学者	国际数字人文组织联盟(ADHO)的奖项委员会SCA负责奖项征集和评选
安东尼奥·赞波利奖(Antonio Zampolli Prize)[5]	2010年至今(每三年)	在重大数字人文项目中有杰出贡献的学者	
数字人文奖(Digital Humanties Awards)	2012年至今	数字人文工具、项目、资源、博客等	志愿者组成的国际题名委员会负责审核和监督，大众通过邮件投票评选
文本数字化成功奖(Succeed Awards)[6]	2013－2014年	数字化的工具	Europeana的专业评审委员会，表彰利用新技术的数字化项目
DHMS(美国中世纪学会数字人文与多媒体研究奖)[7]	2017年至今	美国中世纪研究院内的学者	美国中世纪学会(MAA)颁发
评委会根据MAA对项目的标准选择获奖项目Garfinkel数字人文奖[8]	2016年至今	优秀出版物、项目	美国研究学会(The American Studies Association，ASA)的数字人文小组负责奖项评选
柏林数字人文奖[9]	2015－2018年	数字人文项目	柏林的数字人文会议论坛if｜DH｜b(2013－2018)负责颁奖，会议组委会负责评选
艺术人文杰出成就奖/杰出贡献奖/早期杰出职业奖[10]	2014年至今	加拿大内的机构工作人员或加拿大籍学者	加拿大数字人文学会(CSDH)负责奖项评选，每年轮流颁发
NYCDH研究生奖[11]	2014年至今	纽约州地区的研究生	纽约市数字人文科学小组(NYC)的颁奖委员会负责项目评审
数字人文进步奖(DHAG)[12]	2010年至今	项目资助、奖学金资助	美国国家人文基金会(NEH)负责评审
British Library labs Awards[13]	2015年至今	针对大英图书馆的数据或馆藏开发的项目	大英图书馆的专家和馆员组成的评委会
罗伊·罗森茨威格数字历史创新奖[14]	2009年至今	数字历史的创新项目	由乔治梅森大学的罗森茨威格奖基金会提供资助，美国历史学会(AHA)的专家委员会对项目进行评审

数字人文是一个目标驱动型的研究领域，实践项目丰富，很多学者从项目内容、技术、主题领域等角度总结分析数字人文的现状和需求特征。李慧楠等通过对数字人文年会内容的整理，根据项目对应学科领域进行梳理分类[15]。徐彤阳等总结了澳大利亚数字人文项目概况并建立评价体系[16]。韩豫哲通过对国外数字人文中心网站的项目介绍，总结优秀项目的特点[17]。上述文献的聚类和比较分析以地区或会议论文为样本，缺乏全球不同地区的横向对比、不同时间的纵向分析。因此本文选择了“数字人文奖”的获奖项目作为研究对象，整理归纳获奖项目的内容特征，为数字人文的实践提供启示。“数字人文奖”(DH Awards)由一群数字人文爱好者于2012年发起，目的是为了宣传和表彰数字人文的资源、项目和知识，吸引更多的公众参与数字人文社群，提高数字人文意识。“数字人文奖”的国际学术委员会负责奖项启动和监督投票，成员是来自各国的志愿者。每年三月开始征集题名，国际委员会审核题名资源是否属于数字人文领域、是否本年度发布或取得重大进展，并将资源分配在不同类别下，在为期两周的公开投票期结束后，在其网站上公布每个类别的冠亚季军和其他提名资源的信息[18]。

与ADHO、CSDH等学会颁发的奖项最大的不同是，“数字人文奖”的优胜作品完全由公众投票决定。从“数字人文奖”的获奖项目中，可以归纳优秀案例的共性，从中窥探人文学者心中“好的项目”“实用的工具”和“失败的教训”，为国内数字人文实践提供参考。

2 "数字人文奖"案例论述

"数字人文奖"常设奖项有:"最佳工具"(Best DH Tool or Suite of Tools)、"最佳可视化"(Best DH Data Visualization)、"最佳博文"(Best DH Blog Post or Series of Posts)和"最佳娱乐"(Best Use of DH For Fun)。昙花一现的奖项有:"最佳外语项目"(Best DH contribution not in the English language)只颁发过一届,随着非英语资源的增多,2013 年后便不再以语言区分资源;"最佳资源集"(Best DH Dataset)设立于 2019 - 2020 年,由于参与竞争的资源太多,国际委员会增设了该类别;"最佳失败案例研究"(Best Exploration of DH Failure/Limitations)在 2014 年设立,由于没有足够的资源参与竞选,2018 年起才恢复该类别评选。

2012 年至今,"数字人文奖"累计评选出 162 项获奖作品,2019 年是竞争最激烈的一年,多个资源打成平手,其中,"最佳娱乐"中两个资源并列第一,"最佳博客"和"最佳公众参与"两个资源并列第三;2020 年是提交资源最多的一年,共有 124 项资源参与竞选。从统计结果来看,参与度最高的国家是美国,其次是英国、法国和意大利;近年来,亚太地区,特别是韩国和中国积极参与数字人文实践,2020 年中国提交的资源就有 6 项,可见国内数字人文活动的活跃性。从获奖机构来看,英国的伦敦大学学院(UCL)数字人文中心最为活跃,开发的项目分别获得 2013 年的可视化奖、娱乐奖和 2016 年的可视化奖。

2.1 最佳工具

从数字人文基础设施建设到相关学科讨论,工具一直是数字人文绕不开的话题,甚至有学者认为工具属性是数字人文的根本属性[19],数字人文未来可能会成为一门辅助学科。"最佳工具"奖项的评选是最为连续的,凸显了数字人文对人文学科的支撑和交叉程度。按获奖的工具的特征和类型,大致可分为如下 3 类:

2.1.1 轻量级、开源的编程工具或内容管理工具

这类工具具有开源、易用和技术门槛低的特点,能帮助人文学者快速搭建内容管理平台或构建虚拟学术空间。最典型的轻量级的内容管理系统(CMS)工具是 Omeka[20],具备免费开源、使用简单、社区支持、功能集成且灵活可扩展性的优势,被广泛用于数字人文项目如佛罗里达记忆项目、迈阿密大学的 DLP 项目、布法罗城女性口述史项目等。很多数字人文的获奖项目和平台都是基于 Omeka 搭建,如 2012 年的"最佳公众参与"奖 The Boston Bombing Digital Archive 和 2016 年的"最佳工具"奖 Broken Books。最典型的免费开源的工具包是 2013 年的冠军——纽约市立大学开发的 CBOX[21],是搭建虚拟学术社区和在线学习空间的工具,设置过程直观且支持移动设备的访问,被很多人文学者作为教学工具使用。2019 年的季军比萨大学开发的 EVT(Edition Visualization Technology)是另一款轻量级的开源工具,用于从 XML 编码的文本创建数字版本,可帮助不懂 TEI 的传统人文学者摆脱网络编程的负担[22]。

2.1.2 文本分析类工具

人文计算最早发源于文学和语言学[23],在线词典、语料库、文本标记和翻译等工具是数字人文工具的热点。SHEBANQ(2014 年冠军)是希伯来文的查询工具[24];Dantesource(2015 年冠军)是用 RDF 编码的但丁文集[25];EAGLE(2016 年冠军)是提供古罗马铭文的翻译软件[26];Recogito(2018 年冠军)是文本和图像标注工具[27];CDO(2019 年冠军)是一款在线词典工具,提供古埃及语 - 科普特语与德语、英语、法语互译功能[28];Old English Online(2020 年季军)是帮助初学者提供古英语发声练习的在线工具[29]。

2.1.3 平台类、综合性的服务工具

近年来愈来愈多的工具以平台形式提供综合性服务,ALCIDE(2015 年亚军)是一个基于 Web 的平台,集合了自然语言处理 NPL 的工具包、可视化工具等,从时间、地理和语义角度分析历史文献,方便学者更有效地处理历史文本并分享研究结果[30]。美国 Gale 集团开发的 Digital Scholar Lab(2018 年亚军)是一个基于云研究环境,将 Gale Primary Sources 数据库内文档进行 OCR 识别,并提供文本挖掘和可视化的工具[31]。弗吉利亚理工大学主导的 Civil War Photo Sleuth(CWPS,2018 年季军)平台使用众包机制提高面部识别技术的算法,标注照片中的士兵,并将结果与军事历史档案关联,完善美国内战历史数据[32]。伦敦大学学院开发的 Transkribus(2019 年季军)是一个用于手稿识别和转录的平台,该平台提供的手写识别模型 HTR,准确率达到 95%,支持专家管理个人文件并导出文本,支持众包项目,利用志愿者提高深度学习模型,鼓励志愿者上传文档丰富平台资源[33]。

2.2 最佳可视化

可视化一直是数字人文研究的热点,数字人文学者主要利用图像处理、计算机虚拟环境和用户界面,呈现数据分析结果和项目建设。有学者分析国内外数字

人文研究主题领域的相关研究，发现文本挖掘和可视化技术作为数字人文最常用的研究手段经常联合出现，处于数字人文领域的核心第一象限内[34]。可视化作为通用技术往往穿插在整个项目中，根据使用阶段和呈现效果，获奖的项目可分为以下 4 类型：

2.2.1 虚拟现实重建

利用虚拟现实技术 VR 和增强现实技术 AR 建立虚拟环境，重现历史场景，包括重构历史建筑、模拟历史场景等，构建三维空间的虚拟世界给人文学者以超越时空的“场景再现”能力。2014 年的冠军 Virtual Paul's Cross 项目，利用数字建模技术，虚拟重建了 17 世纪初的圣保罗教堂模型、唱诗班以及教堂墓地周围的建筑，为访问者提供沉浸式的讲道体验。2018 年德国柏林博物馆的 Coins 项目，利用 PixiJS，d3 和 React 将馆藏硬币虚拟展示，并提供多种交互方式[35]。

2.2.2 交互式地图

可视化最广泛的应用和表现形式是将历史史料与地理信息系统（GIS）结合，生成时空交互式地图。2012 年斯坦福大学的 ORBIS 项目，模拟生成古罗马的航运模型，根据用户选择的站点计算路线、时间和成本[36]。2013 年的 Touch History 项目也是交互式地图，用户可以按意愿在 400 多年前的魁北克市进行旅行[37]。2016 年的冠军 Peripleo 项目利用关联数据技术，将不同来源的史料中的地理数据连接，并投放到不同时期的动态地图上，该项目为人文学者提供了开源的历史地理信息的注释工具。2019 年所有的获奖项目都是交互式地图，爱丁堡大学的 Witches 项目将 16 世纪－18 世纪的女巫居住地、审判和死亡地结合时间轴制作了交互式地图；法国国家艺术史研究所和卢浮宫合作，将 19 世纪的古董买卖、价格等信息制作成可视化地图显示 19 世纪法国的文化和社会历史；哈佛大学开发的 The Atlas of Economic Complexity 是一个显示全球地图和贸易数据的交互式地图，成为投资者和学者探索全球贸易动态的重要工具。

2.2.3 社交网络关系可视化

基于人物之间的信息交流构建的社交关系网络，是数字人文研究中的另一个可视化方向，近两年有学者开始将时空和社交网络结合，为人文研究提供更多的视角。最典型的获奖案例是 Kindred Britain（2013 年亚军），将英国 3 万位历史名人通过血缘、婚姻或其他关系进行连接构成网络显示，从时间、地理、网络关系 3 个维度宏观显示英国历史名人[38]。

2.2.4 语料或史料的可视化

文学研究中，人文学者进行“远读”的方式是将语料可视化，对大规模文本内容的可视化主要分为两类：整体特征的描述，如词频统计的词云图；和内在结构特征揭示，如对历史小说的体裁、类型、叙事结构等形成的知识图谱。2015 年冠军 Metaphor Map of English 该项目基于牛津英语词典的历史版本，显示不同领域、不同时期单词不同的涵义。2014 年的 Signs@40 项目是对 *Sings* 期刊 40 年内发表的论文的主题的可视化，展示女权主义的研究主题随时间的变化。2017 年由佐治亚理工学院开发的 The Shape of History 项目，利用网格＋时间轴的方式描述历史事件、颜色表示政治势力或事件类型，将历史事件转化为形状和色彩呈现出来[39]。

2.3 最佳项目/最佳公众参与项目

数字人文是大数据和计算驱动的研究模式，无论是基础设施建设还是数据建设，仅靠单个机构的投入进展缓慢，不仅需要多个文化遗产机构的协作，而且需要公众参与到资源数据的共建共享中，实现数字人文服务的创新。按公众参与的程度和内容，可分为 4 种类型：

2.3.1 转录校正文本

转录校正型是最常见的公众参与项目，主要是手稿、档案、古籍的抄录或转录。德克萨斯理工大学的手稿抄录项目（2013 年季军）、墨西哥国家人类学和历史研究所的门萨法典数字化项目（2015 年亚军）、意大利古文字学项目（2019 年季军）、都是利用众包模式进行的。一些史料的转录甚至需要多学科的专家参与，如 2016 年的季军 Decoding the Civil War 该项目涉及美国内战时期的军用电报和密码，需要具备专业知识的专家和学术进行抄录解读[40]。

2.3.2 资源共建

文化遗产项目资源繁多且分散，需要借助多方力量搜集资源。多机构的合作共建项目如智利记忆（2017），汇集了智利国家图书馆、国家文化遗产部、博物馆等多家机构的藏品。口述历史项目是较为常见的资源建设方式，波士顿马拉松恐怖袭击档案（2013）、美国宗教之声项目（2019）都设有口述历史小组搜集整理档案。另一种方式是号召个人上传资源，Archive Alert（2017）为抢救濒危档案，号召公众上传认为重要的档案或史料。NumET（2018）鼓励学者上传中世纪文化遗产相关资料。

2.3.3 标引数据

数据标引类的项目主要集中在地图标记、文献索

引方面。如 2016 年的 ToposText 项目[41]，召集学生对古希腊文进行索引，并将文本记录与地名词典相连。

2.3.4 协同式知识创造

即公众通过社区互动、开源平台进行分布式、协同式的知识创作。Open a GLAM Lab（2019）是由 16 位来自不同国家的文化遗产机构的专家在 5 天内共同完成一本书；2014 年的 The American Yawp[42] 和 Encyclopedia of the First World War[43] 项目是汇集领域内上百位专家在线集体创作的历史资料，这些项目的实践展示了在虚拟环境下，不同学科的专家依托主题任务进行探讨、解释、表示和知识构建的过程。

2.4 最佳娱乐

数字人文项目可以看作是对数据的创新开发使用，最佳娱乐奖项的设置鼓励人文学者以新颖、创新、有趣的方式使用技术、资源进行探索实践，某些项目可能只是学者的灵光一现，学术参考价值不高，但是对资源、工具的创意使用不仅对其他项目的开发有启示作用，而且可以宣传数字人文。

2015 年冠军 PAGANS 是一个互动性游戏，调查用户对艺术作品相似性判断。该游戏是对意大利的 VVV（Verbo Visuale Virtuale）项目中的艺术作品而开发的，搜集的信息用于分析用户特征和艺术品特征之间的关联性，帮助策展人更科学地建立展览路径[42]。2016 年的冠军 A Generator of Socratic Dialogues 是机器模拟苏格拉底的对话，是机器写作的有趣尝试。2019 年的伦敦国王学院开发的 The Digital Ghost Hunt，将编码教学、增强现实和现场融合在一起，鼓励学生学习编程[44]。

2.5 最佳博客文章或系列文章

在数字学术环境下，人文学者越来越多地通过博客、社区和网站进行交流、互动、分享和发布研究成果，推特、博客等社交网络成为人文学者进行学术传播和交流的重要渠道，Mcpherson 甚至描述数字人文为“博客人文”。获奖博客、网站的作者和研究领域如表 2 所示，从中可以总结出 3 个特征：①个人“单打独斗”到团队合作的转变，早期获奖的博客都是个人学者进行维护和更新，2016 年开始获奖项目所有者大多是团队甚至是多机构的合作，类型也不仅限于博客，包括同行审议的 OA 期刊、数字教程、专题网站等；②各学科的学者关注点明显不同：历史学家是最为活跃的人文学者群，除了开通个人博客发布研究成果和笔记，还合作搜集、编写数字工具、方法、开源程序的教程；以英语专业为代表的语言学家主要关注建立同行审议的 OA 期刊；图书馆、档案馆或高校的数字人文中心主要通过数字平台宣传数字人文活动、资源建设等内容；③团队学科背景由单一到多学科的变化，发布的内容也不再局限于某个学科领域，而是对更为通用、广泛的数字工具和技术的记录和反思。

2.6 最佳失败案例研究

该奖项是“数字人文奖”最为特色的部分，委员会设立奖项的目的是鼓励数字人文学者不畏失败、不断尝试和探索。创新值得嘉奖，带给人文学者学习和启发的失败经历更值得鼓励。学者们可以从失败的实践中吸取经验教训，总结项目中出现的问题并记录解决方案，并给相关从业者带来反思和思考。但是由于提名项目较少，只在 2014、2018、2019、2020 年颁发过 4 届，根据学者的反思和总结，数字人文项目中出现的主要障碍和问题可分为如下 3 类：

2.6.1 版权政策问题

孤儿作品是图书馆等公共文化机构的数字化项目中的资源黑洞，各国对合理使用孤儿作品的法律规章或多或少存在不足。M. Terras 在其博客中记录了如何在英国政策变化不定情况下取得孤儿作品的许可[45]。

2.6.2 与人文学者的交流和团队协作问题

数字人文项目聚集了人文学者、图书馆员、IT 技术人员等多方人员，团队成员能否顺畅沟通和取得共识对项目成效影响巨大。艺术人文的网络技术设施项目 Project Bamboo（2008 – 2012）因未能与人文学者取得紧密联系，而未能构建成服务、社区为导向的体系。

2.6.3 数据保存问题

数字人文项目能否可持续发展，与数据的保存、维护息息相关。C. Barats 等调研了数据收集到传播的所有阶段，认为数据的长期保存不仅考虑软件、工具、技术因素，还要考虑人文研究和实践中的环境因素，并从数据访问、语料库建立、数据分析传播这三方面探讨数据保存方面的矛盾[46]。

3 获奖项目分析和思考

“数字人文奖”的特色可概括为：①参与广泛。奖项的提名不受语言和地理的限制，积极鼓励小语种地区的文化资源参选，参与国遍布全球，是传播度最高的奖项；从 2012 年设立起从未中断，是数字人文领域连续举办时间最长的奖项之一。②公开透明。参与竞选的作品的清单、投票数都公布在网站上，奖项评选过程

表 2　最佳博客获奖列表

年份	名称	所有者	机构	专业/职业	类型
2012	Digital Humanities Now	多人	卢森堡大学	历史	在线出版物
	Leonardo Flores："I ♥ E-Poetry"	L. Flores	阿帕拉契州立大学	英语	学术博客
	'Kafka's Wound'：Re-imagining the Literary Essay for the Digital Age	W. Self，et al		小说家、文学家	文学随笔、评论
2013	'Psychopower' of Cultural Diplomacy in the Information Age	N. Grincheva	牛津大学	博物馆	在线出版物
	Digital Literary Pedagogy：An Experiment in Process-Oriented Pedagogy	R. Whitson，et al	华盛顿州立大学	英语与数字技术	OA 期刊
	The Digital Humanities Is about Breaking Stuff	J. Stommel	玛丽华盛顿大学	英语	OA 期刊
2014	Six Degrees of Spaghetti Monsters				无法访问
	Cork LGBT History	O. Egan		档案学、社会学	数字档案
	The Digital Humanities network on Twitter	M. Grandjean	洛桑大学	历史	学术博客
2015	AWOL：The Ancient World Online Blog	C. E. Jones	纽约大学	历史	开源博客
	Late Antiquity，Digital Humanities，and Musings on the Classical World	S. E. Bond	爱荷华州立大学	历史	学术博客
	Digital Humanities on Twitter，a small-world	M. Grandjean	洛桑大学	历史	学术博客
2016	The Programming Historian	多人		历史	开源教程
	Macro-Etymological Analysis of Milton's Paradise Los	J. Reeve	哥伦比亚大学	计算语言学	学术博客
	Visionary Cross				无法访问
2017	The Programming Historian en español	多人			开源教程
	The Public Medievalist Special Series：Race，Racism and the Middle Ages	多人		历史	OA 期刊
	Colonial and Postcolonial Digital Humanities Roundtable	R. Risam	塞勒姆州立大学	英语	演讲稿
2018	The British Library's Digital Scholarship blog	多人	大英图书馆	图书馆	博客
	OpenMethods	多人	DARIAH-EU	数字人文	博客
	The Digital Orientalist	多人	美国东方学会	数字人文	博客
2019	The Shiloh Project	多人	多个高校	宗教	博客
	01Lab（零壹 Lab）	多人	南京大学	数字人文	微信公众号
	Center for Digital Humanities at Princeton Updates	多人	普林斯顿大学图书馆	数字人文	网站
2020	Ticha Project，2020 Blog Series	多人	哈佛大学	文化遗产	数字浏览器
	Reviews in Digital Humanities	J. Guiliano & R. Risam	印第安纳大学、塞勒姆州立大学	数字人文	OA 期刊
	Mapping the Scottish Reformatio	M. D. Brock & C. R. Langley	华盛顿与李大学（W&L 大学）、纽曼大学	历史学、近现代史	可视化

透明度较高，兼具公平性和民意。从提名到获奖都是公众邮件投票完成的，获奖的唯一依据是选票数，委员会负责筛选有效投票，避免重复投票。③奖项设置灵活全面。涵盖数字人文相关的工具、项目、文章等多个方面，并根据每年竞选资源的数量和内容来调整奖项设置，如 2020 年设置的"新冠专项"（Special Category：Best DH Response to COVID-19），既能体现该奖的创新性和实用性，也能反映当年数字人文的热点。

但是由于没有奖金激励，且没有专业学会的支持，"学院派"对其认可度不高，诸如书信共和国（Republic of Letters）、"威尼斯时间机器"（Venice Time Machine）等知名项目都没有参与其中，可见该奖项更多地是娱乐性和宣传性。从获奖的结果来看，很多有影响力的工具如 Markus、哈佛中国历代人物传记资料库（China Biographical Database，CBDB）、文本编码倡议（Text Encoding Initiative，TEI）等并未获得理想名次，而有些获得名次的项目或工具因为各种原因已停止更新或继续开发，如 Textal（2013 年的"最佳娱乐项目"的亚军）最近更新是在 2014 年，Checklist for Digital Humanities Projects（2017 年"最佳工具"冠军）因资源迁移而无法访问。从评选标准和维度来看，作品能否获奖的唯一指标是网络选票，评价维度较为单一，投票途径只能通过谷歌表单，欧美以外地区特别是国内访问网址和投票有一定障碍，因此获奖结果缺乏一定的专业性和客

观性;项类目设置不统一,奖项的权威性和稳定性也稍显不足。

纵览获奖项目的内容和建设模式,可以发现其具有如下特征:①很多获奖项目的选题角度较小而新颖,但是专业性强,如 2019 年的最佳可视化获奖项目"苏格兰女巫的可视化地图",将女巫的居住地、死亡地以关联数据发布并制作交互式地图。"小而精"的方式受到近年来学者的青睐,特别是对缺少团队支撑、项目处于起步阶段的个人学者而言,这种方式更有助于提高成功率。②社会性议题、资源建设类的项目,借助公众力量成为一种趋势,社会公众与专家、技术人员、高校师生共同参与到数字人文项目的建设、运营和传播,从资料搜集、数据清洗、资金支持到传播分享。③数字人文项目的可持续发展成为重要议题。从调研中发现很多获奖项目已停止运营维护,如在 2012 年、2014 年入选的 Project Bamboo 项目因缺乏人文学者参与、资金短缺等原因而终止。造成数字人文项目失败的原因很多,比如资金短缺、技术过时、人员流动等[47],还有部分早期建设的数据集和项目因资源迁移、数据格式选择错误等原因而无法访问,因此在数字人文项目的立项阶段就要考虑到可持续发展问题,将生命周期管理和后续维护等问题纳入项目规划和建设目标中。

4　总结和建议

4.1　数字人文技术总结

信息技术是人文研究范式革新最重要的驱动,数字人文的技术体系可分为:数字化技术、数据管理技术、数据分析技术、可视化技术、VR 技术和机器学习技术。本文基于获奖的数字人文项目和工具,归纳了目前在数字人文领域应用最广泛的信息技术,如表 3 所示:

表 3　获奖资源核心技术

资源名称	内容	类型	主要技术
Recogito	语义标注工具	工具	IIIF、关联数据、GIS
Transkribus	文档转录平台	工具	AI、机器学习、大数据
Livingstone's 1870 Field Diary	利文斯通冒险日记	数据集	光谱成像技术
Kindred Britain	英国名人家谱	可视化	网络可视化技术
Tudor Networks	书信网络可视化	项目	社交网络分析,可视化
DanteSources	但丁文集	工具	RDF 本体
Newspaper Navigator Dataset	报纸数据集	项目	机器学习
Gale Digital Scholar Lab	数字人文实验室	工具	文本分析、关联数据
DARIAH	艺术人文数字基础设施	项目	关联数据、GIS
COVID19 Cartography	可视化论文	可视化	自然语言处理、算法
Animal Crossing	虚拟空间	可视化	VR

表 3 列举了部分获奖资源和主要技术,资源描述框架(RDF)、本体、关联数据和国际图像互操作框架(International Image Interoperability Framework, IIIF)主要用于数据的组织,文本分析和挖掘技术、光谱成像技术、自然语言处理、社交网络分析算法、机器学习等 AI 技术主要用于数据的处理,地理信息系统(GIS)、虚拟现实(VR)、增强现实(AR)、混合现实(MR)等技术主要用于数据的展示和呈现。有学者认为 RDF、关联数据、IIIF、大数据和 AI 是数字人文项目建设的通用技术框架,RDF、TEI、SGML、IIIF 等通用标准,定义了数据、文档、图像的描述、组织和呈现标准,使得数字人文底层资源实现统一融合;AI 技术、算法模型、社会分析、机器学习等技术用在文本挖掘和图像处理方面,对人文资源进行重塑和改造;越来越多学者借助 GIS、AR 等技术对研究结果进行可视化展示,促进人文知识的传播、理解和创新。当然,也应该看到,技术并非一成不变,数字人文的技术框架也是出于不断发展、更新中,同时,技术也没有好坏之分,在明确项目特点和应用需求后选择合适的技术才是好技术。

4.2　数字人文工具开发建议

人文学者往往是在使用数字工具中接受数字人文并推动数字人文的发展,从某种程度上说,数字人文首先是一种技术工具,在解决传统人文研究问题的过程中,影响人文学者的研究范式和路径。尽管在数字人文基础设施和工具上已经投入了大量资金,但是只有 6% 的人文学者在研究中使用数字资源和工具[33]。有学者指出,数字工具需要更先进的技术、更强大的功能才能吸引人文学者的使用。但从获奖结果来看,最受人文学者好评的工具并非是集成了最先进的技术、最复杂的功能。从获奖工具的共性中,给工具开发商提供如下建议:

4.2.1 轻量开源，社区支持

无论是内容管理工具 OMKEA、可视化工具 EVT 还是标注平台 Recogito，这些工具都是开源免费使用，属于轻量级应用软件，部署方便、技术门槛低。虽然数字人文课程有意识地培养人文学者的数字素养和基本编程能力，但大部分人文学者对编程是零基础的，因此工具的使用需要低门槛，让研究人员在不懂技术知识的前提下也可以操作工具。活跃的用户社区是工具成功运行的另一个重要因素，该社区可以是系统开发交流的平台如依托 Github、Trello 的任务交流，也可以是用户使用互动的平台。开源社区不仅能保证工具版本的更新和功能完善，还可以为人文学者提供技术支持。

4.2.2 界面友好，功能易用

有些工具的界面虽然简陋，但是用户所见即所需，以 Online Coptic Dictionary 为例，其界面简洁，功能显示直观，人文学者甚至依靠直觉就可以使用。近两年工具有往集成化、平台化发展趋势，工具构建者需要在设计和开发中考虑用户界面的灵活性和清晰性，隐藏部分不常用功能，与用户形成高效互动。正如一位历史学家评论说，易用的工具可以让学者更集中研究问题本身，避免在探索功能和操作上浪费时间[48]。

4.2.3 加强工具培训，重视用户文档的编写

很多人文学者对数字工具缺乏了解，包括原理、操作、案例等，因此对工具有“不信任”感，因此有学者呼吁应该重视工具的推广，例如培养、传授对数据的分析、操作、解读等技能。Gale 数字学术实验室在高校内开展培训讲座，因此获得人文学者的好评。F. Gibbs 的实践调研[49]表明，很多人文学者在阅读了工具的文档之后，无法完全理解工具对其研究的帮助，因此建议用户文档包含工具的基础知识、具体实例以及研究方法。文本标注工具 Markus 就是很好的例子，该平台上有详细的操作步骤、演示动画以及案例分享。

4.2.4 以人文学者需求为导向，与人文学者建立密切联系

以史学为例，历史地理是较早引入数字技术的领域，很早引入统计分析软件并建立数据库，但对具体研究问题来说，需要“量身定做”的工具，比如古今地图叠加、度量单位转换、统计数据等。与人文学者建立密切联系，推动我们思考结合研究对象和主题的特性，开发针对性的数字人文工具。

4.2.5 开放数据接口，集成工具平台，为用户提供一站式的服务

随着数字基础设施的完善，数据标准化越来越重要，工具开发者应该和数据存储机构、其他工具合作，鼓励不同格式、标准的兼容，支持数据的互操作性。Markus 从原本的文本标注工具发展为具有多功能的阅读和文本分析平台，集成了多个参考工具和编辑器，支持与 GIS、CBDB 进行数据共享。

4.3 数字人文项目开发建议

数字人文是一个涉及多学科合作的研究领域，具有研究工作量大、经费有限、跨学科合作的特点。众包模式利用开放式参与和群体智慧，进行人文数据的收集、处理和分析，是解决数字人文项目中数据短缺的最优方式。目前数字人文项目的前端建设将越来越多地采用众包模式，促进数字人文基础设施建设。

从失败的项目实践来看，数字人文项目有两个挑战性：项目的可持续和数据的可访问性。大部分项目都存在资金短缺，甚至因为没有后续基金支持而导致网站无法访问。对于资金、资源受限的小型数字人文项目，A. Gil 提出了“最小计算”的项目实践模式，即在硬件、软件、网络容量受限的环境下，以“最优配置”“最小维护量”和“最小消耗量”来构建数字人文项目[50]。具体准则包括：用离散的、轻量级组件来构建项目框架，如采用微服务架构、组件封装、代码模块化等方式来增强程序的灵活性和可扩展性；尽量使用通用的标准定义数据，如使用 W3C 标准、关联数据进行数据发布，增加对资源内容的访问和数据的重用性；降低对工具、数据库和软件的依赖，减少资源需求和处理时间；减少使用专业术语，尽量使用通用语言增加对共同文化的参与，例如领域专家和开发人员一起创建通用语言，通过团队交流简单清晰地传递业务规则，增强项目的可维护性。

5 结语

通过纵向比较“数字人文奖”的获奖资源，可以回顾数字人文的应用脉络：①从工具角度来看，早期数字人文应用聚焦于文本内容管理、文本编码等文本处理方面，学科交叉不断深入，加入了地理、时空等维度的内容标注，随着数字人文方法论和学术共同体的的构建完善，以 Transkribus 为代表的软件不仅仅是一个工具，同时也为学者构建了一个虚拟和开放的研究环境甚至是知识生产系统。②从学科应用领域来看，数字人文的学科应用随时间的推移不断深入，语言学和历史学未来仍是数字人文应用的主要学科，艺术、宗教神学、女权主义哲学、后殖民主义等社会人文学科的应用频次也在逐渐提升，同时学科的交叉和融合的特征更

为明显,学者间的跨界交流合作加强,这种变化趋势在博客文章中尤为明显。③从数字资源角度来看,特色资源和“冷门绝学”的研究更容易获得奖项青睐,如对彩虹群体的研究、小众语言或文化的资源集、后殖民主义的思考等,说明只要找准题材、定位,具有文化传承和研究价值,“小项目”也能获奖。通过对“数字人文奖”获奖资源的研究,本文对数字人文的工具、项目进行梳理和分析,并对数字人文工具的开发、项目建设提出建议,在未来的研究中,还需要对数字人文的实践进行更广泛的收集和分析,以探索数字人文具体发展状况。

参考文献:

[1] 陈苗,刘晗月. 数字人文研究热点与发展趋势的断面考察——以《数字人文季刊》和《人文学科中的数字计算》为中心[J]. 图书馆研究与工作,2020(1):10-16.
[2] 柯平,宫平. 数字人文研究演化路径与热点领域分析[J]. 中国图书馆学报,2016,42(6):13-30.
[3] 贝里,费格约德. 数字人文:数字时代的知识与批判[M]. 王晓光,等译. 大连:东北财经大学出版社,2019.
[4] ADHO 奖[EB/OL]. [2021-02-15]. https://adho.org/awards.
[5] 安东尼奥赞波利奖[EB/OL]. [2021-02-15]. https://eadh.org/awards/antonio-zampolli-prize.
[6] 文本数字化成功奖[EB/OL]. [2021-02-15]. https://www.succeed-project.eu/.
[7] 美国中世纪学会数字人文与多媒体研究奖 DHMS[EB/OL]. [2021-02-15]. https://www.medievalacademy.org/page/DHPrize.
[8] Garfinkel 数字人文奖[EB/OL]. [2021-02-15]. https://www.theasa.net/communities/caucuses/digital-humanities-caucus/garfinkel-prize-digital-humanities.
[9] 柏林数字人文奖 Berliner DH-Preis[EB/OL]. [2021-02-15]. https://www.ifdhberlin.de/dh-preis/.
[10] 加拿大数字人文奖[EB/OL]. [2021-02-15]. https://csdh-schn.org/awards/.
[11] NYCDH 研究生奖[EB/OL]. [2021-02-15]. https://nycdh.org/nyc-dh-graduate-student-project-award/.
[12] 数字人文进步奖 DHAG[EB/OL]. [2021-02-15]. www.neh.gov/grants/odh/digital-humanities-advancement-grants.
[13] 大英图书馆实验室奖[EB/OL]. [2021-02-15]. https://data.bl.uk/bl_labs_awards.
[14] 罗伊罗森茨威格数字历史创新奖[EB/OL]. [2021-02-15]. https://www.historians.org/awards-and-grants/awards-and-prizes/roy-rosenzweig-prize.
[15] 李慧楠,王晓光. 数字人文的研究现状——“2019 数字人文年会”综述[J]. 情报资料工作,2020,41(4):49-59.
[16] 徐彤阳,杨明睿. 澳大利亚数字人文项目透视——兼论对我国数字人文项目的启示[J]. 图书情报工作,2020,64(22):145-152.
[17] 韩豫哲. 国外数字人文项目特点分析及启示[J]. 图书馆,2019(11):18-23.
[18] Digital humanities awards[EB/OL]. [2021-03-01]. http://dhawards.org/.
[19] 张正. 数字人文的工具属性研究[J]. 图书馆研究,2019,49(5):19-23.
[20] 邱建华. Omeka 系统在数字人文研究中的应用剖析[J]. 情报探索,2019(10):104-109.
[21] Commons in a box [EB/OL]. [2021-02-01]. http://commonsinabox.org/.
[22] EVT[EB/OL]. [2021-02-15]. http://evt.labcd.unipi.it.
[23] 王晓光. “数字人文”的产生、发展与前沿[C]//教育部人文社会科学研究方法创新论坛论文集. 武汉:武汉大学出版社,2010:207-220.
[24] SHEBANQ[EB/OL]. [2021-02-10]. http://shebanq.ancient-data.org/.
[25] DanteSources[EB/OL]. [2021-02-13]. http://perunaenciclopediadantescadigitale.eu/dantesources/index.html.
[26] EAGLE[EB/OL]. [2021-01-23]. https://wiki.eagle-network.eu/wiki/index.php/Main_Page.
[27] Recogito[EB/OL]. [2021-02-10]. https://openpublishingawards.org/index.php/recogito/.
[28] CDO[EB/OL]. [2021-02-10]. http://coptic-dictionary.org/search.cgi.
[29] Old English online[EB/OL]. [2021-02-13]. https://oldenglish.info/about.html.
[30] Alcide[EB/OL]. [2021-02-13]. http://celct.fbk.eu:8080/Alcide_Demo/.
[31] Digital scholar lab[EB/OL]. [2021-02-18]. https://www.gale.com/intl/primary-sources/digital-scholar-lab.
[32] Civil war photo sleuth[EB/OL]. [2021-02-20]. http://www.civilwarphotosleuth.com.
[33] Transkribus[EB/OL]. [2021-02-12]. https://read.transkribus.eu/transkribus/.
[34] 颜石磊. 国内外数字人文研究主题比较分析[J]. 大学图书情报学刊,2021,39(1):107-111.
[35] The virtual paul's cross project[EB/OL]. [2021-02-15]. http://vpcp.chass.ncsu.edu/.
[36] Orbis[EB/OL]. [2021-02-22]. http://orbis.stanford.edu/.
[37] Touch history[EB/OL]. [2021-02-15]. https://www.youtube.com/watch?v=_l8Vjb-87uE.
[38] Kindred britain[EB/OL]. [2021-02-15]. http://kindred.stanford.edu/.
[39] The shape of history[EB/OL]. [2021-02-25]. http://shapeofhistory.net.
[40] Decoding the civil war[EB/OL]. [2021-02-15]. https://www.zooniverse.org/projects/zooniverse/decoding-the-civil-war.

[41] ToposText [EB/OL]. [2021-02-15]. https://topostext.org.

[42] Open a glam lab[EB/OL]. [2021-02-23]. https://glamlabs.io/books/open-a-glam-lab/.

[43] Verbo Visuale Virtuale[EB/OL]. [2021-02-23]. http://dh.fbk.eu/projects/vvv-verbo-visuale-virtuale-la-piattaforma-di-ricerca-interattiva-dellarte-verbo-visuale.

[44] Digital ghost hunt[EB/OL]. [2021-02-23]. https://www.kdl.kcl.ac.uk/our-work/digital-ghost-hunt/.

[45] Reuse of digitised content[EB/OL]. [2021-02-20]. http://melissaterras.blogspot.co.uk/2014/10/reuse-of-digitised-content-4-chasing.html.

[46] Quinn Dombrowski. What ever happened to project bamboo? [EB/OL]. [2020-12-25]. http://llc.oxfordjournals.org/content/29/3/326.

[47] Towards a taxonomy of failure[EB/OL]. [2021-02-25]. http://quinndombrowski.com/blog/2019/01/30/towards-taxonomy-failure.

[48] IATH. Summit on digital tools for the humanities[R]. Charlottesville: University of Virginia, 2006.

[49] Fred Gibbs. Building better digital humanities yools: toward broader audiences and user-centered designs[J]. Digital humanities, 2012,6(2):1-14.

[50] The user, the learner and the machines we make [EB/OL]. [2021-02-15]. http://go-dh.github.io/mincomp/thoughts/2015/05/21/user-vs-learner/.

作者贡献说明：

单蓉蓉：文献调研、论文撰写和修改；

刘炜：确定论文选题，设计论文框架，提出修改意见；

陈涛：论文修改；

李惠：论文修改，为论文提供素材和指导。

Features and Suggestions for the Development of Digital Humanities Projects: Based On the Discussion and Analysis of the International Digital Humanities Award Winning Projects

Shan Rongrong[1] Liu Wei[2] Chen Tao[3] Li Hui[4]

[1] Department of Library, Information and Archives, Shanghai University, Shanghai 200444

[2] Shanghai Library, Shanghai Institute of Science and Technology Information, Shanghai 200031

[3] School of Information Management, Sun Yat-sen University, Guangzhou 510006

[4] College of Humanities & Social Development, Nanjing Agricultural University, Nanjing 210032

Abstract: [**Purpose/significance**] In the field of international digital humanities, Digital Humanities Award is one of the most authoritative awards. Through combing and analyzing the award-winning projects, the commonalities between "excellent projects" and "practical tools" in the field of digital humanities can be found, and the shortcomings of failed projects can be summarized. [**Method/process**] This paper used literature research method, content analysis method and information measurement method to sort out the award-winning projects of Digital Humanities Award over the years, and summarized the characteristics and deficiencies of the award-winning projects. [**Result/conclusion**] After investigation and analysis, this article proposes that the digital humanities tool development should be oriented towards code open source and user needs; project practice should be based on linked data, international image interoperability framework (IIIF), resource description framework (RDF) and artificial intelligence technology (AI) as the main technical framework; these projects with limited funds or resources can choose "minimum computing" as the practice mode.

Keywords: Digital Humanities Awards tool development technology realization minimal computing